DOCUMENT RESUME

ED 133 353 TM 005 965

AUTHOR Stake, Robert E.
TITLE Overview and Critique of Existing Evaluation Practices and Some New Leads for the Future.
PUB DATE [Apr 76]
NOTE 4p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Evaluation; *Evaluation Methods; Evaluation Needs

ABSTRACT

This paper initiates the AERA symposium on issues and methods in qualitative evaluation with a theoretical overview and critique of the state of the evaluation field in the United States. Some of the basic theoretical underpinnings to a qualitative approach to evaluation are presented. This type of approach complements and compensates for those aspects of educational practice that conventional approaches omit. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED133353

OVERVIEW AND CRITIQUE OF EXISTING EVALUATION PRACTICES AND SOME NEW LEADS FOR THE FUTURE*

Robert E. Stake
Center for Instructional Research and Curriculum Evaluation
University of Illinois at Urbana-Champaign

We live in a cruel and greatly tolerant society. It is cruel in that few persons, few programs, and few institutions are allowed to become more than our expectations of them. The society is tolerant of all sorts of variety and offers millions of ways of being successful.

Few people demand any justification of that which already exists. We have a great rhetoric of accountability, competition, payment for results, and "the proof is in the pudding." But those in privileged positions avoid scrutiny. And only the innovative are called up to demonstrate their effectiveness.

Evaluators are called to the scene: agents of rationality, planned change, and enlightenment. Theirs is a position of privilege. They speak of social service, technical precision, and utilitarianism. They work hard; spell well; and protect their client, the funding agency, and their research colleagues. They ask complex, though seldom embarrassing, questions. They are of what Basil Bernstein calls the new middle class, controlling the words that control the machines that keep society orderly and productive. They defer to merit and efficiency as they see it.

Some of my best friends are evaluators. They are good people, believing in all sincerity that theirs is a helping profession, that they are inching toward truth--if only people would pay attention and learn how to interpret the data.

For all its complexity, educational evaluation is a very ordinary cog in the social, economic, politicized machinery. Here are some current points of contact.

ITEM. The big RFPs for evaluation continue to expect that the evaluation methods will be those of social-survey research and experimentation even though these methods deliberately attempt to avoid subjective judgment and valuing, even though these methods are intent upon contributing to the understanding of education in general rather than to the understanding of the particular programs studied or to particular programs like them.

ITEM. Harrassed by criticism of early efforts to present uninterpreted data, National Assessment outgoing director Stan Ahmann interpreted the most recent decline of science scores to mean that science teaching in America was becoming poorer, a conclusion which may be true but one for which he had neither empirical nor experiential grounds. The validity of the National Assessment tests has not been established for "state-of-the-health-of-education" purposes or for any policy-setting purposes. Furthermore, no standardized tests have ever been shown to have statistically based validity for making decisions about programs at the district, state, or federal level.

TM005 965

---

*Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, April 1976.

ITEM. With some two-thirds of the states having mandated some form of state assessment--following vigorous efforts in Michigan and Florida for half a dozen years--the plunge into this type of state control of education has leveled off, mostly because the early results were not found to be useful, were costly, and were even becoming politically embarrassing.

ITEM. Thousands of districts find themselves forced to increase substantially their budgets for data management to meet the appetite for reports from top people in the district superintendent's office or in the state superintendent's office, thus further depleting moneys for curricular programs in a time of rising costs and fixed school incomes.

ITEM. Recently Lovington High School dropped its membership in the North Central Association, giving up its accreditation, and the University of Illinois College of Education withdrew from AACTE membership, thus losing its NCATE accreditation. Institutional accreditation, based on self-study and observations of visiting peers, is increasingly challenged as cumbersome and costly. Too many faculty members see the self-study as a time to hoodwink administrators, site visitors, and the public, none of whom are thought to understand the programs anyway. But, mainly, the self-study is disappointing because it doesn't help to satisfy state or federal data requirements.

ITEM. Trainers of evaluation specialists mobilize for expanded work. Their estimates of need for new manpower numbering into the thousands. College departments are creating new courses; organizations such as AERA expand their evaluation-training-institute offerings; new organizations such as the Evaluation Network and the Evaluation Research Society of America are aborning, as are new periodicals such as the Journal for the Study of Educational Evaluation, the Faculty Development and Evaluation Newspaper, and CAP Newsletter. Still many signs of a growth industry--but most of it depends on what the governments require.

ITEM. Oregon congresswoman Edith Greene, once a leader in the development of federal education programs, is a supporter no longer, largely because of the large proportion of funds skimmed off for administration, research, development, and evaluation. Countless little semicommercial companies and consultancies have joined the biggies: AIR, Rand, SDC, HUMRO, ETS, ABT Associates to compete for the dollars. Still, most of the middlemen are school- and university-based specialists.

ITEM. In 1964, David Ausubel wrote about the necessity of medical-education evaluators knowing the content and issues of medical education. Almost no one paid attention, then or now. When qualifications for evaluators are enumerated, it is considered nice--but not essential--if the evaluator has a working knowledge of the subject matter. Psychologists are encouraged to define and evaluate the basic skills in reading; statisticians are encouraged to analyze Title I programs. Cross-fertilization is not without value, of course; but most evaluators are only semiliterate in the fields they are evaluating, particularly in Education itself. They can write objectives, develop tests, design experiments, and read computer feedback; but they cannot speak intelligently of pedagogy, curriculum, epistemology, or school-community relations.

ITEM. The National Science Foundation is assiduously searching for new program review and evaluation procedures. Previous evaluation projects did not alert them to adverse public reaction to the NSF curriculum "Man a Course of Study." Almost none of their evaluators have been questioning whether the original funding decisions made sense, and few evaluation reports have called for negative action. NSF has a hard time explaining what it has been getting for its evaluation dollar.

The evaluation business may be a growth industry, but certainly one with growing pains. Too much has been expected, too much has been promised. Formalistic evaluation studies have seldom yielded information worth its cost. At best they educate a few evaluators who may then make recommendations that are helpful. But you can't count on it.

The key mistake, I think, is the assumption that objective information can be aggregated across large numbers of teachers or students to provide a basis for decision-making to people who are not personally acquainted with the programs.

The key hope, I think, is that subjective information, based on key issues, oriented to real problems in particular situations, rigorously cross-examined, will become a standard offering of evaluation studies.

It seems to me to be a matter of epistemology: How do we know our programs? What counts as evidence that a program is working? On many occasions I have suggested that the best practical knowledge is direct personal experience and that the most reasonable thing an evaluator can offer is vicarious experience through portrayals and case studies. Such data are particularly valuable if the decisions about programs are to be made in the individual classroom.

It is too much to hope that any information can improve the ability of a distant superintendent or commissioner to tell a teacher what to do or a purchasing agent what to purchase. Still, the expanding distribution of authority and the increase in avowals of responsibility coming from district, state, and federal offices require that we keep trying.

It is not reasonable to suppose that an improved evaluation technology will make education more effective or solve society's problems. Quantitative technocracy is not working in the public interest, but it could be less of a bind on the small business of getting school taught and helping kids get educated.

We wonder if there are nonquantitative techniques which would lead to a more realistic knowledge of programs without inflating the costs, the red tape, and the social constraints of evaluation.